CLAIMS

What is claimed is:

1. An apparatus for automatically extracting metadata from electronic documents

5   comprising a first processing element, a second processing element, a reasoning element,

and a database, wherein,

   i)   said first processing element is further configured to convert electronic

         documents into files;

   ii)   said first processing element is configured to provide the files to a second

10   processing element;

   iii) said second processing element is configured to receive said files and

         extract predetermined information;

   iv)   said second processing element is further configured to provide said extracted

         predetermined information to said reasoning element;

15   v)   said database is configured to also provide input to said reasoning element;

   vi)   said reasoning element is configured to use a set of rules to extract metadata

         from the files; and

   vii) said reasoning element provides an output of metadata.

20   2. An apparatus for automatically extracting metadata from electronic documents as set

forth in claim 1, wherein said files are substantially format invariant data files such as

Postscript files

3. An apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein said predetermined information is substantially spatial layout facts.

5

4. An apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein the second processing element and said database simultaneously input to the reasoning element.

10   5. An apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein said set of rules can be updated.

6. An apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein said metadata is substantially comprised of title, author,

15   affiliation, author affiliation, and table of contents.

7. An apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein said metadata is provided to a user interface.

20   8. An apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein said metadata is provided to a storage medium.

9. A method for automatically extracting metadata from electronic documents providing a first processing element, a second processing element, a reasoning element, and a database and comprising the steps of:

    a) using said first processing element to convert electronic documents to files;

    b) further using said first processing element to provide the files to said second processing element;

    c) using said second processing element to receive said files and extract predetermined information;

    d) further using said second processing element to provide extracted predetermined information to said reasoning element;

    e) using said database to provide input to said reasoning element;

    f) using a set of rules in said reasoning element to extract metadata from the files;

    g) providing an out put of metadata from said reasoning element.

10. The method for automatically extracting metadata from electronic documents as set forth in claim 9, wherein said files are substantially format invariant data files such as Postscript files.

11. A method for automatically extracting metadata from electronic documents as set forth in claim 9, wherein said predetermined information is substantially spatial layout facts.

Method for Automatically Extracting Metadata
From Electronic Documents Using Spatial Rules
HRL065/PD#000211

12. A method for automatically extracting metadata from electronic documents as set forth in claim **9**, wherein the second processing element and the database simultaneously input to the reasoning element.

5   13. A method for automatically extracting metadata from electronic documents as set forth in claim **9**, wherein said set of rules can be updated.

14. A method for automatically extracting metadata from electronic documents as set forth in claim **9**, wherein said metadata is substantially comprised of title, author,

10   affiliation, author affiliation, and table of contents.

15. A method for automatically extracting metadata from electronic documents as set forth in claim **9**, wherein said metadata is provided to a user interface.

15   16. A method for automatically extracting metadata from electronic documents as set forth in claim **9**, wherein said metadata is provided to a storage medium.

Method for Automatically Extracting Metadata
From Electronic Documents Using Spatial Rules
HRL065/PD#000211